



Diethel, T., & Flach, P. A. (2015). *Smart-Homes for eHealth: Uncertainty Management and Calibration*. Paper presented at NIPS 2015 Workshop on Machine Learning in Healthcare, Montreal, Canada.

Peer reviewed version

License (if available):  
Unspecified

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Neural Information Processing Systems at <http://papers.nips.cc/>.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

---

# Smart-Homes for eHealth: Uncertainty Management and Calibration

---

**Tom Diethe, Peter Flach**  
Intelligent Systems Laboratory  
University of Bristol  
Bristol BS8 1UB, UK  
{tom.diethe, peter.flach}@bristol.ac.uk

## Abstract

We introduce the SPHERE project which is devoted to eHealth in a smart-home context, and discuss the implications for such a system in terms of the quantification and management of uncertainty for automated decision making in health care. We then discuss the importance of calibration in such systems, particularly in light of the differing operational contexts that will be encountered.

## 1 Shifting to a new health-care paradigm: the SPHERE IRC

Due to well-known demographic challenges, traditional regimes of health-care are in need of re-examination. Many countries are experiencing the effects of an ageing population, which coupled with a rise in chronic health conditions is expediting a shift towards the management of a wide variety of health related issues in the home. In this context, advances in Ambient Assisted Living (AAL) are providing resources to improve the experience of patients, as well as informing necessary interventions from relatives, carers and health-care professionals.

To this end the EPSRC-funded “Sensor Platform for HEalthcare in a Residential Environment (SPHERE)” Interdisciplinary Research Collaboration (IRC) [13, 14] has designed a multi-modal system driven by data analytics requirements. The system is under test in a single house, and will be deployed in a general population of 100 homes in Bristol (UK). The data sets collected will be made available to researchers in a variety of communities.

Current research on data fusion and machine learning in the SPHERE project addresses two main challenges, which are transparent decision making under uncertainty and adapting to multiple operating contexts. We proceed to describe these two challenges.

## 2 First challenge: Transparent decision making under uncertainty

Naturally, the SPHERE setting presents many sources of uncertainty. Firstly, we are dealing with multiple sensor modalities (environmental, body-worn, video), each of which will have different noise profiles and failure modes. Secondly, we are dealing with a situation where annotated or labelled data is expensive and intrusive to acquire, and the resulting labels are potentially noisy and inaccurate (indeed in some cases there may be no “ground truth” in the classical sense, and we need to resort to modelling annotator disagreement explicitly). Lastly, patterns of human behaviour are subject to many factors (internal and external) that may or may not be attributed to the particular health context of a given individual.

Faced with such a situation, the most sensible approach would be to use “white-box” modelling methods as far as possible. Model-based machine learning [5, 12] attempts to follow this ideal by encoding assumptions about the problem domain explicitly in the form of a model. Indeed,

the model can be viewed simply as this set of assumptions, expressed in a precise mathematical form. These assumptions include the number and types of variables in the problem domain, which variables affect each other, and what the effect of changing one variable is on another variable. The result is that any decisions made by the system can be inspected, so that if the model is performing poorly, the solution is to re-examine the assumptions being made.

In the Bayesian paradigm, degrees of belief in states of nature are specified through the use of probabilities, which through the construction of probabilistic graphical models [10] allow us to apply a principled mathematical framework of the quantification of uncertainty to perform model-based machine learning. On the basis of the models we build, Bayesian decision theory tries to quantify the trade-off between various decisions, making use of probabilities and costs [4, 3].

A typical problem that we face is that the differences between individuals will too large to be captured by a single model. Hierarchical Bayesian models [7] allow us to simultaneously generalise over communities of residents whilst also learning personalised models. In addition, they allow us to be more flexible with our priors, by specifying “hyper-priors”, and then performing inference over the priors instead.

### 3 Second challenge: Adapting to multiple operating contexts

However, transparently dealing with degrees of belief does not solve all modelling challenges posed by the SPHERE project. Our models and inferences have to be applied in multiple contexts, and indeed any given context is liable to both gradual and abrupt shifts. In such situations, it will be crucial that we are able to trust the probabilities coming from the system. A machine learning system is well “calibrated” if the predicted probabilities it gives correspond to observed frequencies. This is natural in forecasting (we would expect it to rain in 60% of cases where a weather forecaster predicts a 60% chance of rain [11]) but carries over to machine learning as well. If a system is poorly calibrated then it suggests a problem either in the model (such as an overly restrictive assumption) or in the inference.

Different operating contexts also call for different performance metrics which perhaps incorporate a different notion of expected loss [8]. If the goal is to minimise loss, for example for the case of classification, a systematic approach would be that given a model, threshold choice methods that correspond with the available information about the operating condition should be applied, followed by comparison of their expected losses. Different classification performance metrics such as F-score also imply a different notion of calibration [6]. More generally, the choice of the performance metrics in use should be seen as another modelling assumption rather than being independent from the model. Given that we expect the end users of our systems to include medical professionals as well as the residents themselves, we can easily see how the types of decision we would want to surface should be adaptable.

Explicitly modelling context change also favours domain adaptation and model reuse. We are building on the results of the REFRAME project [1], which developed a general methodology for model reuse in machine learning called *reframing* [9]. The setting is exemplified by the recent ECML-PKDD’15 Discovery Challenge *MoReBikeS: Model Reuse with Bike rental Station data* [2], which encouraged participants to build predictive models for new bicycle rental stations making use of previously trained models on other stations (for which the training data was however no longer available).

### Acknowledgements

This work was performed under the SPHERE IRC funded by the UK Engineering and Physical Sciences Research Council (EPSRC), Grant EP/K031910/1. The project is actively working towards releasing high-quality data sets to encourage community participation in tackling the issues outlined here.

### References

- [1] <http://reframe-d2k.org/>.
- [2] <http://reframe-d2k.org/Challenge>.

- [3] James O Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.
- [4] José M Bernardo and Adrian FM Smith. *Bayesian theory*, volume 405. John Wiley & Sons, 2009.
- [5] Christopher M. Bishop. Model-based machine learning. *Phil Trans R Soc A*, 371, 2013.
- [6] Peter A Flach and Meelis Kull. Precision-Recall-Gain curves: PR analysis done right. In *Proceedings of the Twenty-Ninth Annual Conference on Neural Information Processing Systems*. NIPS, 2015.
- [7] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*, volume 2. Taylor & Francis, 2014.
- [8] José Hernández-Orallo, Peter Flach, and Cèsar Ferri. A unified view of performance metrics: Translating threshold choice into expected classification loss. *J. Mach. Learn. Res.*, 13(1):2813–2869, October 2012.
- [9] José Hernández-Orallo, Ricardo B.C. Prudêncio, Meelis Kull, Peter Flach, Chowdhury Farhan Ahmed, Nicolas Lachiche, and Adolfo Martínez-Usó. Reframing in context: A methodology for model reuse in machine learning. 2015. Submitted.
- [10] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [11] Allan H Murphy and Robert L Winkler. Probability forecasting in meteorology. *Journal of the American Statistical Association*, 79(387):489–500, 1984.
- [12] John Winn, Christopher M. Bishop, and Tom R. Diethe. *Model-Based Machine Learning*. Microsoft Research, 2015.
- [13] Przemyslaw Woznowski, Xenofon Fafoutis, Terence Song, Sion Hannuna, Massimo Camplani, Lili Tao, Adeline Paiement, Evangelos Mellios, Mo Haghighi, Ni Zhu, et al. A multi-modal sensor infrastructure for healthcare in a residential environment. In *Communication Workshop (ICCW), 2015 IEEE International Conference on*, pages 271–277. IEEE, 2015.
- [14] Ni Zhu, Tom Diethe, Massimo Camplani, Lili Tao, Alison Burrows, Niall Twomey, Dritan Kaleshi, Majid Mirmehdi, Peter A. Flach, and Ian Craddock. Bridging e-health and the internet of things: The SPHERE project. *IEEE Intelligent Systems*, 30(4):39–46, 2015.